



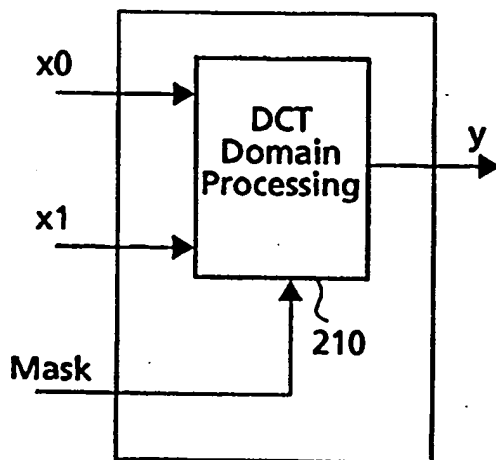
## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>H04N 7/30, 5/265</b>		A1	(11) International Publication Number: <b>WO 99/18735</b>
			(43) International Publication Date: 15 April 1999 (15.04.99)
(21) International Application Number: PCT/US98/20783 (22) International Filing Date: 2 October 1998 (02.10.98) (30) Priority Data: 60/061,246      7 October 1997 (07.10.97)      US 09/003,273      6 January 1998 (06.01.98)      US (71) Applicant: THOMSON CONSUMER ELECTRONICS [US/US]; 10330 N. Meridian Street, Indianapolis, IN 46290-1024 (US). (72) Inventor: JONSSON, Ragnar, H.; Ibud 203, Breidavik 39, IS-112 Reykjavik (IS). (74) Agent: CARTER, Jeffrey, D.; Thomson Consumer Electronics, 10331 North Meridian Street, Indianapolis, IN 46290-1024 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  Published With international search report.	

(54) Title: PICTURE MASKING AND COMPOSITING IN THE FREQUENCY DOMAIN

## (57) Abstract

Processing image signals. At least one image signal and a mask signal are received, wherein the image signal and mask signal are in the discrete cosine transform (DCT) domain. Masking of the image signal is performed in the DCT domain, in accordance with the mask signal, by representing the masking in terms of the DCT basis functions, to provide an output image signal.

200

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LJ	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## PICTURE MASKING AND COMPOSITING IN THE FREQUENCY DOMAIN

5

### BACKGROUND OF THE INVENTION

#### Field of the Invention

The present invention relates to video processing systems, and, in particular, to apparatuses and methods for performing picture masking and compositing in the DCT domain.

#### Description of the Related Art

10 Computer systems are frequently used to perform various types of video or image processing, such as picture masking and compositing. In masking, a specified fraction of certain pixels of a first image are retained in a new image, according to a provided mask. In compositing, pixels of two input images are combined or "blended" at a certain ratio, to form a new image.

Such masking and compositing are important operations, for example in commercial video  
15 or image processing. For example, commonly used effects such as chroma keying, wipe, and overlaying are based on compositing pictures from two video sources. Masking and compositing are also frequently used in production of still images, for example, when generating magazine advertisements and posters.

Computer systems are also used for various data encoding purposes, such as video  
20 compression. Many video compression standards (including JPEG, MPEG-1, MPEG-2, H.261, and H.263) are based on the discrete cosine transform (DCT), it may be desirable to process compressed pictures in the DCT domain. However, image processing techniques like masking and compositing are typically designed to operate in the spatial domain, not the frequency, or DCT, domain. Thus, if the image processing of compressed video signals is done in the spatial domain, the input  
25 compressed video signals must be transformed into the spatial domain before being processed, and the processed signal must be transformed back into the DCT domain once more. Such transformation to the spatial domain and back into the frequency domain can be very computationally expensive and, therefore, undesirable. Moreover, conventional "brute force" convolutions performed directly in the frequency domain are also extremely computationally  
30 expensive.

### **SUMMARY**

In the present invention, at least one image signal and a mask signal are received, wherein the image signal and mask signal are in the DCT domain. Masking of the image signal is performed in the DCT domain, in accordance with the mask signal, by representing the masking in terms of the DCT basis functions, to provide an output image signal.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

Fig. 1 shows a prior art spatial domain image processing system;

Fig. 2 is a block diagram of a DCT domain image processing system, in accordance with a preferred embodiment of the present invention; and

Fig. 3 depicts an exemplary processed image processed by the DCT domain image processing system of Fig. 2.

### DESCRIPTION OF THE PREFERRED EMBODIMENT

As explained above, performing image processing techniques like masking and compositing in the spatial domain offers drawbacks when processing compressed video signals, which are in a frequency domain such as the DCT domain. Accordingly, in the present invention, there is provided an efficient method and associated apparatus for implementing picture masking and compositing in the DCT domain. As described in further detail below, the technique of the present invention is based on representing the masking function in terms of the DCT basis functions and computing the masking as a weighted sum of the results of masking by the DCT basis functions.

In the DCT domain, masking by the DCT basis functions has a relatively simple and efficient implementation. Because of the energy compaction property of the DCT, the weight of many of the functions is very small and can be dropped from the weighted sum without introducing noticeable artifacts. This leads to very efficient implementations for masking and compositing images in the DCT domain, typically requiring less than three multiplications per pixel. These and other features and advantages of the present invention are described in further detail below.

#### Spatial Domain Processing of DCT Images

Referring now to Fig. 1, there is shown a prior art spatial domain image processing system 100. As illustrated, spatial domain image processing system 100 includes three inverse DCT (IDCT) functional blocks 120, 121, 122, and a DCT functional block 130, as well as spatial domain processing functional block 110. As will be appreciated by those skilled in the art, each of these functional blocks may be implemented in hardware or software. For example, the IDCT and DCT operations of blocks 120, 121, 122, and 130, respectively, as well as the spatial domain processing of block 110, may be performed by a suitably programmed general-purpose or special-purpose microprocessor.

System 100 receives as input signals the mask signal and image signals  $x_0$  and  $x_1$ , each of which are in the DCT domain. For example, image signals  $x_0$  and  $x_1$  may have been previously compressed with a process that utilizes the DCT. System 100 outputs output image signal  $y$ , which represents the compositing of image signals  $x_0$  and  $x_1$  in accordance with the mask signal. Output image signal  $y$  is also in the DCT domain. Since block 110 performs image processing in the spatial domain (e.g., with RGB or YUV spatial representations of image pixels), IDCT blocks 120, 121, and 123 are necessary in prior art systems to transform the input signals into the spatial domain.

Once the (spatial domain) input signals are processed, the processed output signal must be transformed back into the DCT domain, to provide signal  $y$ .

As will be appreciated, it is trivial for spatial domain processing unit 110 to implement spatial masking in the spatial domain by using spatial windowing. For an input picture  $x[m,n]$ ,  
 5 masking (also referred to as windowing) with the mask, or window,  $w[m,n]$ , is simply

$$y[m,n] = w[m,n] \times x[m,n] \quad (1)$$

As will be appreciated, windowing in the spatial domain is equivalent to convolution in the frequency domain. The masking in (1) can, therefore, be implemented by DCT processing of DCT signals as

$$Y[k,l] = W[k,l] * X[k,l] \quad (2)$$

where  $X[k,l]$ ,  $Y[k,l]$ , and  $W[k,l]$  are the frequency representations of  $x[m,n]$ ,  $y[m,n]$ , and  $w[m,n]$ ,  
 10 respectively,  $*$  is the convolution operator,  $m, n$  are the spatial domain indices, and  $k, l$  are the DCT or frequency domain indices. The approach in (2) is a "brute force" DCT domain processing implementation based on symmetric convolution. As will be appreciated, a symmetric convolution is achieved by making a symmetric extension of two finite length signals and the convolving the extended signals together using circular convolution. If the frequency domain is the discrete  
 15 Fourier transform (DFT) domain, the convolution in (2) is circular convolution. Further background on such techniques may be found in D.E. Dudgeon & R.M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1984. If the frequency domain is the DCT domain (or other discrete trigonometric transforms), the convolution in (2) is a symmetric convolution. Further background on symmetric convolutions may be found  
 20 in S.A. Martucci, *Symmetric Convolution and the Discrete Sine and Cosine Transforms: Principles and Applications*, PhD thesis, Georgia Institute of Technology, 1993. Spatial masking in the DCT domain can, therefore, be implemented by using symmetric convolution according to (2).

Masking can be used to implement compositing of two input pictures  $x_0[n,m]$  and  $x_1[n,m]$  according to

$$y[m,n] = x_1[m,n] + w[m,n] \times (x_0[m,n] - x_1[m,n]) \quad (3)$$

In this case a mask value of one means that samples from  $x_0[n,m]$  are used, while a mask value of zero means that samples from  $x_1[n,m]$  are used. Mask values in the range from zero to one imply linear interpolation between the two signals  $x_0$  and  $x_1$ . (Mask values outside the unit interval imply linear extrapolation of the two input pictures.)

5           If the mask,  $w[m,n]$ , is a separable signal, the convolution in (2) can be implemented as two separate one-dimensional (1-D) convolutions. As will be appreciated, a 2-D signal  $x[m,n]$  is separable if there exist two 1-D signals  $r[m]$  and  $s[n]$  such that  $x[m,n] = r[m]s[n]$  (i.e., it can be implemented as a cascade of horizontal and vertical DCTs). In the separable case, the convolution may provide a reasonable approach to masking, since it requires, for example, only 16  
10       multiplications per sample for an 8x8 DCT. For non-separable masks, however, the convolution approach to masking is not as feasible since, for example, masking for an 8x8 block DCT requires 64 multiplications and considerable data shuffling.

          Accordingly, both spatial domain processing of DCT images and the brute force DCT approach often require an undesirably high amount of computation. As video compression resulting  
15       in DCT images becomes more common, it becomes more desirable to do picture processing on compressed image data without completely decoding or decompressing the image data. Some techniques in this regard are discussed in further detail in B.C. Smith & L.A. Rowe, "Algorithms for Manipulating Compressed Images," *IEEE Computer Graphics & Applications*, pp. 34-42, September 1993; S-F Chang & D.G. Messerschmitt, "A New Approach to Decoding and  
20       Compositing Motion-Compensated DCT-Based Images," *ICASSP-93*, pp. V421-V424, 1993; and N. Merhav & V. Bhaskaran, "A Transform Domain Approach to Spatial Domain Image Scaling," *ICASSP-96*, pp. 2405-2409, 1996.

### **Frequency Domain Processing of DCT Images**

25       In the present invention, compressed pictures are processed in the DCT domain with a technique based on representing the masking function in terms of the DCT basis functions and computing the masking as a weighted sum of the results of masking by the DCT basis functions, as described in further detail below. Such DCT domain processing makes it possible to reduce both the computational complexity and the latency of the processing, by eliminating the need for transforming signals from the DCT domain into the spatial domain and back.

As will be appreciated, if the desired processing (i.e., masking and compositing) is done in the DCT domain, the three IDCT transforms 120, 121, 122, and the DCT transform 130, required in spatial domain processing of DCT images can be eliminated. Referring now to Fig. 2, there is shown a block diagram of a DCT domain image processing system 200, in accordance with a preferred embodiment of the present invention. As shown, system 200 comprises DCT domain processor 210, but does not comprise nor require the three IDCT transforms and one DCT transform used in spatial domain processing. Instead, DCT domain processor 210 operates in the DCT domain, and is thus able to provide processing efficiencies relative to spatial domain processing.

In one embodiment, system 200 operates with respect to two-dimensional (2-D) type-II DCT of 8x8 blocks, such as is used by the image and video compression standards JPEG, MPEG-1, MPEG-2, H.261 and H.263. As will be appreciated, however, in alternative embodiments the present invention may be utilized with other types of DCTs and other block sizes.

The 8x8 type-II DCT is given by

$$X[k,l] = \eta[k]\eta[l] \sum_{m=0}^7 \sum_{n=0}^7 x[m,n] \cos\left(\frac{\pi}{16}(2m+1)k\right) \cos\left(\frac{\pi}{16}(2n+1)l\right) \quad (4)$$

where  $\eta$  is a frequency-dependent DCT normalization coefficient which depends on the values of DCT domain indices  $k, l$ . It should be noted that even though the 2-D DCT can be used to *represent* non-separable signals, the transform itself is separable—and the basis functions of the 2-D DCT are separable. DCT basis functions are discussed in further detail below, with reference to (15).

The 2-D DCT of each block can be implemented using matrix multiplications

$$X = C\bar{X}C^T, \quad (5)$$

where  $X$  and  $\bar{X}$  are matrix representations of  $X[k,l]$  and  $x[m,n]$ , respectively, and  $C$  is the DCT transformation matrix (a unitary matrix, i.e.,  $CC^T=I$ ). For the 8x8 DCT, the matrices  $\bar{X}$ ,  $X$  and  $C$  are all 8x8.

Vertical spatial masking by any window  $v_v[m]$  and horizontal masking by any window  $v_h[n]$  can be implemented as the matrix multiplication



$$\bar{Y}_{k,l} = \bar{V}_k \bar{X} \bar{V}_l \quad (6)$$

where

$$\bar{V}_k = \text{diag}(v_k[n]) \quad (7)$$

Based on the fact that C is unitary, it can be derived that

$$Y_{k,l} = C \bar{Y}_{k,l} C^T \quad (8)$$

$$= C \bar{V}_k \bar{X} \bar{V}_l C^T \quad (9)$$

$$= C \bar{V}_k C^T C \bar{X} C^T C \bar{V}_l C^T \quad (10)$$

$$= V_k X V_l \quad (11)$$

where

$$V_j = C \bar{V}_j C^T. \quad (12)$$

5 It should be noted that non-separable masking cannot be expressed in a simple matrix multiplication form similar to (11). However, a non-separable mask can be transformed by the 2-DCT, which does have separable basis functions. The IDCT of the DCT domain representation of the non-separable mask,  $W[k,l]$ , is given by

$$v[m,n] = \sum_{k=0}^7 \sum_{l=0}^7 W[k,l] \eta[k] \cos\left(\frac{\pi}{16}(2m+1)k\right) \eta[l] \cos\left(\frac{\pi}{16}(2n+1)l\right) \quad (13)$$

Substituting (13) into (1) gives

$$y[m,n] = \sum_{k=0}^7 \sum_{l=0}^7 \{ (W[k,l] \eta[k] \eta[l]) v_k[m] x[m,n] v_l[n] \} \quad (14)$$

where

$$10 \quad v_k[m] = \cos\left(\frac{\pi}{16}(2m+1)k\right) \quad (15)$$

This implies that masking  $x[m,n]$  with  $w[m,n]$  is equivalent to a weighted sum of the masking of  $x[m,n]$  with the basis functions of the IDCT. It should be noted that the windowing in (14) is separable and can, therefore, be written in terms of the matrix multiplications in (6), where both  $v_k[m]$  and  $v_l[n]$  are given by (15) for  $k,l \in \{0,1,\dots,7\}$ . Therefore, the DCT transform of the masked signal becomes

$$Y = \sum_{k=0}^7 \sum_{l=0}^7 \{(W[k,l] \eta[k]\eta[l]) V_k X V_l\} \quad (16)$$

where the numerical values of the DCT domain windowing matrices,  $V_j$ , can be evaluated according to (12).

Thus, a non-separable mask can be implemented as weighted sum of separable functions (13), and masking can be accomplished with separable functions using matrix notations (11). Therefore, non-separable masks can be implemented as a weighted sum of separable masking operations (16). The separable masking operations in (16), by the DCT basis functions, then turn out to have simple and efficient implementation.

As will be appreciated, the functions defined in (15) are the DCT basis functions (for 1-D type-II DCT of size  $N=8$ ). The DCT basis functions form an orthogonal basis that can represent all discrete functions of length  $N$ . The factor  $\eta$  normalizes the basis functions so  $\eta[k]$  times the basis function in (15) (i.e.,  $\eta[k]v_k[m]$ ) forms an orthonormal (normalized orthogonal) basis for all functions of length  $N$ . Since the basis functions for the 2-D DCT are formed as the product of two 1-D basis functions  $v_k[m]$ ,  $v_l[n]$ , the 2-D DCT basis functions are separable.

The windowing matrices for the DCT basis functions,  $V_j$ , are sparse and have very regular structure. For example, for  $j = 4$  we have

$$V_4 = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ \sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{bmatrix}, \quad (17)$$

and the other windowing matrices have the same kind of structure.

By incorporating the factor of  $\frac{1}{2}$  into the windowing function  $W[k,l]$ , each matrix multiplication in (16) can be implemented using only one addition per sample and two multiplications (by  $\sqrt{2}$ ) per 64 samples (for  $8 \times 8$  DCT). If the DCT coefficients are obtained from decoding JPEG or MPEG streams, the multiplications by  $\sqrt{2}$  can be incorporated into the quantization matrices, reducing the computational complexity to only one addition per sample for each matrix multiplication in (16). In addition, there is one multiplication and one addition per pixel for each term in the weighted sum in (16). Therefore, the computational complexity of implementing masking according to (16) is approximately one multiplication and three additions per pixel for each term that is evaluated. Additionally, when the weighting coefficient,  $W[k,l]$ , is zero, the whole term can be dropped and no computation is needed for that term.

As will be appreciated, the DCT approach is used in compression systems, such as JPEG and MPEG, because for most signals the energy is concentrated into relatively few DCT coefficients. In the present invention, this property is utilized to save computations by skipping all processing for weighting coefficients,  $W[k,l]$ , equal to zero. Alternatively, the savings can be made more substantial by dropping weighting coefficients close to zero. As will be appreciated, when the weight is zero or close to zero, terms can be dropped from the sum, which reduces the computational complexity. That is, the representation of the masking in terms of the weighted sum allows computational complexity to be reduced by skipping all processing for weighting coefficients  $W[k,l]$  equal to zero (or, in one embodiment, for all weighting coefficients  $W[k,l]$  less than a predetermined threshold).

As will be appreciated, by adjusting the threshold for choosing which coefficients are dropped, the quality of the masking operation can be traded for computational complexity in a similar manner as quality is traded for bit rate in encoding. The ability to trade off quality of the masking against computational complexity gives great flexibility in trading cost for quality.

5 Accordingly, the frequency domain implementation of picture masking and compositing of the present invention can be very efficient.

Thus, as will be appreciated by those skilled in the art, in the present invention, the masking function is implemented in terms of the DCT basis functions. As will be appreciated, any necessary scaling is first performed, and may be incorporated into the quantization matrix in an  
10 inverse quantization. Next, a weighted sum of the blocks masked in this fashion is then implemented. In one embodiment, the masked block at this point is re-normalized, in accordance with the scaling done previously. (As will be appreciated, the initial scaling and re-normalization scaling may be incorporated into the quantization matrix if the input signal is dequantized and the output signal is quantized.)

15 In one embodiment, all processing for weighting coefficients  $W[k,l]$  equal to zero is skipped (where, for an input picture  $x[m,n]$ ,  $w[m,n]$  is the window used to mask the input picture, and  $W[k,l]$  is the frequency representation of  $w[m,n]$ ). In an alternative embodiment, because of the aforementioned energy compaction property of the DCT, all processing is skipped for weighting coefficients  $W[k,l]$  close to zero. In general, all processing is skipped for weighting coefficients  
20  $W[k,l]$  less than or equal a predetermined threshold value (where a threshold value of zero yields the former case). In one embodiment of the present invention, this threshold is selected in accordance with a desired tradeoff between the quality of the masking operation and computational complexity, where a higher threshold provides lower quality but greater savings in computational complexity, and vice-versa.

25 Compositing of two images can be implemented by use of masking, according to (3).

For a given block of the image (e.g., an 8-by-8 block), for example, the following steps may be taken by a suitably programmed processor to implement the present invention, in one embodiment. First, examine every DCT coefficient of the mask,  $W[k,l]$ , and if the coefficient is  
30 "relevant" (i.e., either bigger than zero or bigger than a given threshold, depending on the embodiment), then do masking by the corresponding basis function, multiply each coefficient of the

masked signal by the weighting coefficient, and add the result to the weighted sum for the block. This implements the weighted sum in (16).

5 The masking by the DCT basis functions can be implemented in terms of matrix multiplications as shown in (11) (and (16)). However, a more efficient implementation can be achieved by taking into account the regular structure of the windowing matrices as the example in (17) shows. Several of these more efficient implementations are discussed in the detailed description above.

10 For processing of original signals already in the DCT domain, the frequency domain processing of the present invention requires less computation than both spatial domain processing and brute force DCT domain processing based on symmetric convolution. Through empirical testing and modeling, the inventors have found that the computational complexity involved in using the frequency domain processing of the present invention is approximately one to four multiplications per sample for most typical masking operations. As will be appreciated, by using an algorithm similar to rate control algorithms, the complexity of spatial masking in the DCT domain can be limited to only three multiplications per sample without any noticeable degradation of the masking quality.

20 By contrast, a single 2-D DCT takes about three multiplications per sample, and when implementing masking of JPEG or MPEG compressed pictures in the spatial domain, IDCTs must be first used to transform the DCT data into the spatial domain, and then use the DCT operation to transform the processed picture back into the DCT or frequency domain. Thus, when implementing picture compositing, there are at least two IDCTs (one for each input picture) and one DCT (for the composite picture) needed, in addition to the spatial processing. Therefore, there are at least nine multiplications needed for implementing picture compositing in the spatial domain, which is approximately three times what is needed for described embodiments of the DCT domain implementation of the present invention. In sum, therefore, for picture compositing of pictures in the DCT domain, the present invention, in one embodiment, requires about three times fewer multiplications per pixel than spatial domain processing, and about twenty times fewer multiplications than processing based on brute force convolution.

25 Referring now to Fig. 3, there is depicted an exemplary image 300 processed using DCT domain picture compositing performed in the frequency domain by image processing system 200.

Image 300 contains a head-and-shoulder portion 312, which is overlaid over a flower garden background 310, and a transparent logo "SARNOFF" 315, which was inserted in the top right hand corner of image 300. The picture compositing performed by system 200 to arrive at image 300 was performed, in one actual experiment, using only 1.8 multiplications per pixel.

5 As will be appreciated, although the embodiments of the present invention described above is implemented with respect to the DCT frequency domain, the present invention is also potentially applicable to other frequency domains in which the masking function may be represented in terms of the frequency domain's basis functions and in which the masking can then be computed as a weighted sum of the results of masking by these basis functions. For example, the present invention  
10 may be applicable to other frequency domains such as the DFT and discrete sine transform (DST).

As will be understood, the present invention can be embodied in the form of computer-implemented processes and apparatuses for practicing those processes. The present invention can also be embodied in the form of computer program code embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other computer-readable storage medium, wherein,  
15 when the computer program code is loaded into and executed by a computer, the computer becomes an apparatus for practicing the invention. The present invention can also be embodied in the form of computer program code, for example, whether stored in a storage medium, loaded into and/or executed by a computer, or transmitted over some transmission medium, such as over electrical wiring or cabling, through fiber optics, or via electromagnetic radiation, wherein, when the  
20 computer program code is loaded into and executed by a computer, the computer becomes an apparatus for practicing the invention. When implemented on a general-purpose microprocessor, the computer program code segments configure the microprocessor to create specific logic circuits.

It will be understood that various changes in the details, materials, and arrangements of the parts which have been described and illustrated above in order to explain the nature of this  
25 invention may be made by those skilled in the art without departing from the principle and scope of the invention as recited in the following claims.

CLAIMS

What is claimed is:

- 1           1. A method for processing image signals, comprising the steps of:  
2           (a) receiving at least one image signal and a mask signal, wherein the image signal and  
3           mask signal are in a frequency domain; and  
4           (b) performing masking of the image signal in the frequency domain, in accordance with the  
5           mask signal, by representing the masking in terms of the basis functions of the  
6           frequency domain, to provide an output image signal.
- 1           2. The method of claim 1, wherein the frequency domain is the discrete cosine transform  
2           (DCT) domain.
- 1           3. The method of claim 2, wherein step (b) comprises the steps of:  
2           (1) masking blocks of the image signal with the DCT basis functions to provide masked  
3           blocks; and  
4           (2) performing a weighted sum of the masked blocks.
- 1           4. The method of claim 2, wherein the image signal is divided into 8x8 blocks and the DCT  
2           is a two-dimensional type-II DCT.
- 1           5. The method of claim 1, wherein step (a) comprises the step of receiving first and second  
2           image signals, the method further comprising the step of compositing the first and second image  
3           signals in accordance with the mask signal.
- 1           6. The method of claim 1, wherein step (b) comprises the step of skipping all processing for  
2           weighting coefficients  $W[k,l]$  equal to zero, wherein the image signal is represented by  $x[m,n]$ , the  
3           mask signal is represented by a window  $w[m,n]$ , and  $W[k,l]$  is the frequency representation of  
4            $w[m,n]$ .
- 1           7. The method of claim 1, wherein step (b) comprises the step of skipping all processing for  
2           weighting coefficients  $W[k,i]$  less than or equal to a specified threshold, wherein the image signal

3 is represented by  $x[m,n]$ , the mask signal is represented by a window  $w[m,n]$ , and  $W[k,l]$  is the  
4 frequency representation of  $w[m,n]$ .

1 8. The method of claim 7, further comprising the step of selecting the threshold in  
2 accordance with a desired tradeoff between the quality of the masking of step (b) and the  
3 computational complexity required to perform the masking of step (b), wherein a higher threshold  
4 provides lower masking quality but smaller computational complexity, and vice-versa.

5 9. An apparatus for processing image signals, the apparatus comprising:

6 (a) means for receiving at least one image signal and a mask signal, wherein the image  
7 signal and mask signal are in a frequency domain; and

8 (b) means for performing masking of the image signal in the frequency domain, in  
9 accordance with the mask signal, by representing the masking in terms of the basis  
10 functions of the frequency domain, to provide an output image signal.

1 10. The apparatus of claim 9, wherein the frequency domain is the discrete cosine transform  
2 (DCT) domain.

1 11. The apparatus of claim 10, wherein means (b) comprises:

2 (1) means for masking blocks of the image signal with the DCT basis functions to provide  
3 masked blocks; and

4 (2) means for performing a weighted sum of the masked blocks.

5 12. The apparatus of claim 10, wherein the image signal is divided into 8x8 blocks and the  
6 DCT is a two-dimensional type-II DCT.

7 13. The apparatus of claim 9, wherein means (a) comprises means for receiving first and  
8 second image signals, the apparatus further comprising means for compositing the first and second  
9 image signals in accordance with the mask signal.

1 14. The apparatus of claim 9, wherein means (b) comprises means for skipping all  
2 processing for weighting coefficients  $W[k,l]$  equal to zero, wherein the image signal is represented



3 by  $x[m,n]$ , the mask signal is represented by a window  $w[m,n]$ , and  $W[k,l]$  is the frequency  
4 representation of  $w[m,n]$ .

5 15. The apparatus of claim 9, wherein means (b) comprises means for skipping all  
6 processing for weighting coefficients  $W[k,l]$  less than or equal to a specified threshold, wherein the  
7 image signal is represented by  $x[m,n]$ , the mask signal is represented by a window  $w[m,n]$ , and  
8  $W[k,l]$  is the frequency representation of  $w[m,n]$ .

1 16. The apparatus of claim 15, further comprising means for selecting the threshold in  
2 accordance with a desired tradeoff between the quality of the masking of step (b) and the  
3 computational complexity required to perform the masking of step (b), wherein a higher threshold  
4 provides lower masking quality but smaller computational complexity, and vice-versa.

1 17. A storage medium having stored thereon a plurality of instructions for processing image  
2 signals, wherein the plurality of instructions, when executed by a processor, cause the processor to  
3 perform the steps of:

- 4 (a) receiving at least one image signal and a mask signal, wherein the image signal and  
5 mask signal are in a frequency domain; and  
6 (b) performing masking of the image signal in the frequency domain, in accordance with the  
7 mask signal, by representing the masking in terms of the basis functions of the  
8 frequency domain, to provide an output image signal.

1 18. The storage medium of claim 17, wherein:

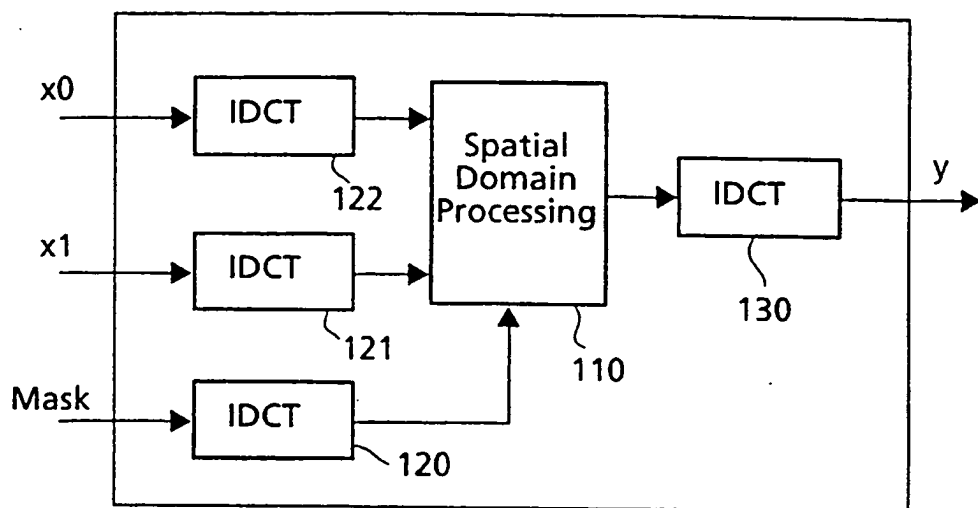
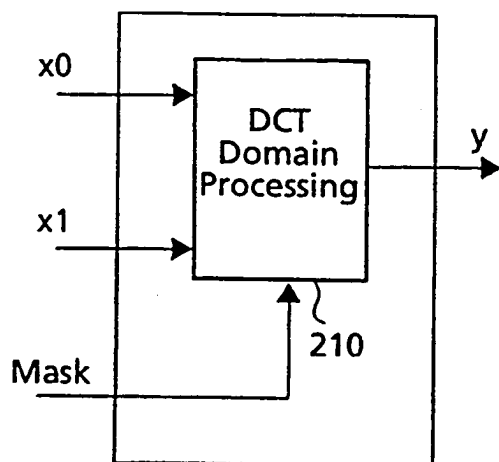
2 the frequency domain is the discrete cosine transform (DCT) domain; and  
3 step (b) comprises the steps of:

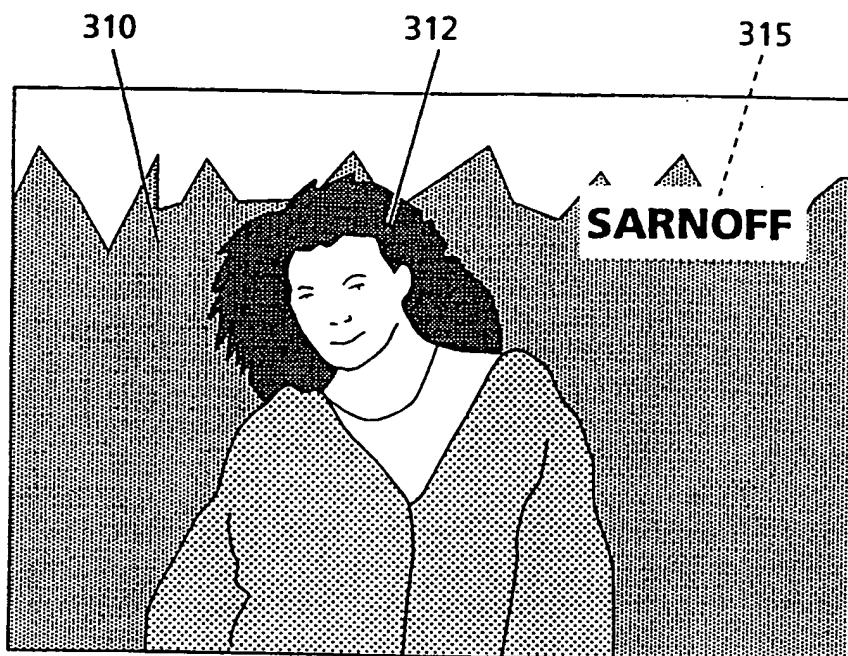
- 4 (1) masking blocks of the image signal with the DCT basis functions to provide  
5 masked blocks; and  
6 (2) performing a weighted sum of the masked blocks.

1 19. The storage medium of claim 17, wherein step (b) comprises the step of skipping all  
2 processing for weighting coefficients  $W[k,l]$  less than or equal to a specified threshold, wherein the

3 image signal is represented by  $x[m,n]$ , the mask signal is represented by a window  $w[m,n]$ , and  
4  $W[k,l]$  is the frequency representation of  $w[m,n]$ .

1 20. The storage medium of claim 19, further comprising the step of selecting the threshold  
2 in accordance with a desired tradeoff between the quality of the masking of step (b) and the  
3 computational complexity required to perform the masking of step (b), wherein a higher threshold  
4 provides lower masking quality but smaller computational complexity, and vice-versa.

**Fig.1**100**Fig.2**200



**Fig.3** 300

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/20783

A. CLASSIFICATION OF SUBJECT MATTER  
 IPC 6 H04N7/30 H04N5/265

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 21079 A (UNIV CALIFORNIA) 15 September 1994	1,2,4-6, 9,10,12, 13,17
Y	see page 18 - page 24	7,8, 14-16, 19,20
A		3,11,18
Y	SMITH B C ET AL: "COMPRESSED DOMAIN PROCESSING OF JPEG-ENCODED IMAGES" REAL-TIME IMAGING, vol. 2, no. 1, February 1996, pages 3-17, XP000656168 see abstract see page 7 - page 9	7,8, 14-16, 19,20
	---	
	-/-	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

11 January 1999

Date of mailing of the international search report

18/01/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

Foglia, P

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/20783

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SMITH B C ET AL: "ALGORITHMS FOR MANIPULATING COMPRESSED IMAGES" IEEE COMPUTER GRAPHICS AND APPLICATIONS, vol. 13, no. 5, 1 September 1993, pages 34-42, XP000562744 cited in the application see the whole document ---	1-20
X	WO 95 33342 A (ICTV INC) 7 December 1995  see the whole document ---	1,2,4,5, 9,10,12, 13,17
X	EP 0 595 218 A (NIPPON ELECTRIC CO) 4 May 1994 see abstract -----	1,2,4,9, 10,12,17

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 98/20783

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9421079 A	15-09-1994	US 5408274 A AU 6395594 A	18-04-1995 26-09-1994
WO 9533342 A	07-12-1995	US 5594507 A EP 0761066 A JP 10505470 T	14-01-1997 12-03-1997 26-05-1998
EP 0595218 A	04-05-1994	JP 6141185 A DE 69319539 D KR 128737 B US 5565925 A	20-05-1994 13-08-1998 04-04-1998 15-10-1996